

On the 3D structure of the mass, metallicity, and SFR space for SF galaxies

Maritza A. Lara-López

mlopez@aao.gov.au

Australian Astronomical Observatory, PO Box 296, Epping, NSW, 1710, Australia

Ángel R. López-Sánchez

Australian Astronomical Observatory, PO Box 296, Epping, NSW, 1710, Australia, Department of Physics and Astronomy, Macquarie University, NSW 2109, Australia

Andrew M. Hopkins

Australian Astronomical Observatory, PO Box 296, Epping, NSW, 1710, Australia

ABSTRACT

We demonstrate that the space formed by the star-formation rate (SFR), gas-phase metallicity (Z), and stellar mass (M_\star), can be reduced to a plane, as first proposed by Lara-López et al. We study three different approaches to find the best representation of this 3D space, using a principal component analysis, a regression fit, and binning of the data. The PCA shows that this 3D space can be adequately represented in only 2 dimensions, i.e., a plane. We find that the plane that minimises the χ^2 for all variables, and hence provides the best representation of the data, corresponds to a regression fit to the stellar mass as a function of SFR and Z , $M_\star = f(Z, \text{SFR})$. We find that the distribution resulting from the median values in bins for our data gives the highest χ^2 . We also show that the empirical calibrations to the oxygen abundance used to derive the Fundamental Metallicity Relation (Nagao et al.) have important limitations, which contribute to the apparent inconsistencies. The main problem is that these empirical calibrations do not consider the ionization degree of the gas. Furthermore, the use of the N_2 index to estimate oxygen abundances cannot be applied for $12 + \log(\text{O}/\text{H}) \gtrsim 8.8$ because of the saturation of the $[\text{N II}] \lambda 6584$ line in the high-metallicity regime. Finally we provide an update of the Fundamental Plane derived by Lara-López et al.

Subject headings: Galaxies: abundances — Galaxies: evolution — Galaxies: fundamental parameters — Galaxies: star formation

1. Introduction

Stellar mass (M_\star), metallicity (Z), and star-formation rate (SFR) are key galaxy properties. M_\star reflects the amount of gas locked up in stars over a galaxy's history. SFR indicates the current rate at which gas is being converted into stars. Z reflects the gas reprocessed by stars over the course of stellar evolution, and any exchange of gas between the galaxy and the environment. The relationships between these three properties are fundamental in understanding galaxy evolution. Indeed, models of galaxy formation within the Λ -CDM scenario already include chemical hy-

drodynamic simulations (e.g., De Lucia et al. 2004; Tissera et al. 2005; De Rossi et al. 2006; Davé & Oppenheimer 2007; Martínez-Serrano et al. 2008).

In the last few years it has been found that M_\star , Z , and SFR are strongly interrelated. Analyzing galaxy measurements from the Sloan Digital Sky Survey (SDSS), Ellison et al. (2008) found that the mass-metallicity (M_\star - Z) relation for star-forming (SF) galaxies depends on the SFR. Subsequently, Lara-López et al. (2010a) reported the existence of a Fundamental Plane (FP) between these three parameters. These authors confirmed that the M_\star - Z and M_\star -SFR relations are just particular cases of a more

general relationship. Lara-López et al. (2010a) fitted a plane and derived an expression for the stellar mass as a function of the gas metallicity and SFR (the Fundamental Plane, FP). In a parallel and independent study, using the same SDSS data, but different Z and SFR estimations, Mannucci et al. (2010) found a similar fundamental relationship, but instead expressed Z as a combination of M_\star and SFR with a substantially different quantitative relationship. They refer to this correlation as the Fundamental Metallicity Relation (FMR). In a recent study, Yates et al. (2012) used models and SDSS data to analyze the dependences of different combinations between SFR, Z and M_\star . They found that the fit given by Mannucci et al. (2010) does not significantly reduce the dispersion of the metallicity compared to the M_\star – Z relation.

A fundamental requirement in all these analyses is obtaining a reliable estimation of the galaxy metallicity. The most robust method to derive the metallicity in SF galaxies is via the estimate of metal abundances and abundance ratios, in particular through the determination of the gas-phase oxygen abundance. This is typically achieved through the analysis of emission-line spectra of H II regions within the galaxies. A proper determination of the oxygen abundance relies on the detection of the [O III] $\lambda 4363$ auroral line (the T_e method, e.g., López-Sánchez & Esteban 2009), but this emission line is usually not observed because of its faintness. Consequently, it is common to invoke the so-called strong-line methods. These techniques assume that the oxygen abundance of an H II region can be derived using only a few bright emission lines. Empirical calibrations based on photoionization models, however, systematically over-predict by 0.2–0.6 dex the oxygen abundances derived using the T_e method and those calibrations which are based on it (see Yin et al 2007; Kewley & Ellison 2008; Bresolin et al. 2009; López-Sánchez & Esteban 2010; Moustakas et al. 2010; López-Sánchez et al. 2012). We discuss these issues further in § 3 below.

Here we explore three different approaches to the representation of the three-dimensional distribution of M_\star , SFR, and Z for galaxies. We detail our sample selection in § 2 and review some issues with metallicity estimators in § 3. The analysis is presented in § 4, and we explore the implications for relationships between SFR and Z in § 5. We present a discussion of the outcome of our analysis, and summarise our results, in § 6.

2. Sample selection

We use data from the Sloan Digital Sky Survey Data Release 7 (SDSS–DR7, Abazajian et al. 2009; Adelman-McCarthy et al. 2007), using the emission-line analysis performed by the MPA-JHU group¹. From the SDSS-DR7 database, we selected galaxies in the “main galaxy sample” (Strauss et al. 2002), with apparent Petrosian r magnitude of $14.5 < m_r < 17.77$ and in the redshift range $0.04 < z < 0.33$. The lower limit ensures that at least 20% of the galaxy light will be inside the 3'' of the SDSS fiber, which is the minimum required to avoid aperture effects (Kewley et al. 2005). To ensure reliable metallicities, we imposed a minimum signal-to-noise ratio, $\text{SNR} > 8$, for each of the most prominent lines H α , H β , [N II] $\lambda 6584$, [O III] $\lambda 5007$, and [O II] $\lambda 3728$. The specific SNR threshold chosen turns out not to be critical to the results. The impact of SNR choice, and the selection of lines it is imposed upon, is explored in detail by Foster et al. (in prep.), who show that for SNRs between about 3 and 8, the resulting mass-metallicity relation does not change substantially. For higher SNR thresholds, though, the weaker (often lower SNR) lines may be excluded, leading to a reduced sensitivity to the high-metallicity population.

We selected SF galaxies following the criteria of Kauffmann et al. (2003) on the Baldwin, Phillips & Terlevich (1981) diagram, $\log([\text{O III}] \lambda 5007 / \text{H}\beta)$ vs. $\log([\text{N II}] \lambda 6584 / \text{H}\alpha)$. The above criteria give us a sample of 45 475 galaxies.

We use gas-phase metallicities, total SFRs, and stellar masses derived by the MPA-JHU group. Metallicities were estimated through a Bayesian approach based on simultaneous fits to all the most prominent lines according to Tremonti et al. (2004, hereafter, T04), while total SFRs were obtained from Brinchmann et al. (2004). Aperture effects are again an important issue to consider in the estimate of SFR (Brinchmann et al. 2004). Yates et al. (2012) shows that if SFRs are calculated without making an aperture correction, nearby galaxies will have their SFRs underestimated, and this leads to a reduced spread in the SFR distribution for low mass, low redshift galaxies. This will have an impact on any exploration of the SFR dependence of the M_\star – Z relation. Both Tremonti et al. (2004) and Brinchmann et al. (2004) make use of the population synthesis and photoionization codes given by Charlot & Longhetti (2001). Stel-

¹<http://www.mpa-garching.mpg.de/SDSS>

lar masses were estimated from fits to the photometry as described in Kauffmann et al. (2003).

3. Metallicity estimate issues

Since the calculation of metallicity is a particularly challenging process, the most accurate and reliable measurements available should be used whenever possible. Approximating metallicity estimates through simple parameterisations can be a valuable tool when only limited information is at hand, but such approximations have significant limitations and uncertainties. For example, the $N_2 \equiv \log([N II] \lambda 6584 / H\alpha)$ and $R_{23} \equiv ([O II] \lambda 3727 + [O III] \lambda \lambda 4959, 5007) / H\beta$ parameters can be used to estimate metallicity with the empirical calibrations of Nagao, Maiolino & Marconi (2006, hereafter, N06). The N06 calibrations are cubic fits between the R_{23} or the N_2 parameter and the T04-derived metallicity (see Fig. 1).

It is well known however (e.g. Baldwin, Phillips & Terlevich 1981; McGaugh 1991; Kewley & Dopita 2002; Pilyugin & Thuan 2005; López-Sánchez & Esteban 2010; López-Sánchez et al. 2011) that the degree of ionization of the gas plays a fundamental role in deriving a reliable estimation of the oxygen abundance. The most common empirical calibrations based on either a direct determination of the T_e (Pilyugin & Thuan 2005; Pilyugin et al. 2010) or photoionization models (Kewley & Dopita 2002; Kobulnicky & Kewley 2004; Tremonti et al. 2004) do consider an ionization parameter. Without such information, the uncertainty of the derived oxygen abundances may be as large as 0.25 – 0.40 dex (López-Sánchez & Esteban 2010).

Indeed, even the T_e method itself is subject to uncertainties, as temperature fluctuations and gradients can render the T_e method incorrect by up to 0.4 dex (e.g., López-Sánchez et al. 2007; Peimbert et al. 2007; García-Rojas & Esteban 2007; Esteban et al. 2009; Mesa-Delgado & Esteban 2010), and absolute metallicity estimates are challenging to make at best (see recent review by López-Sánchez et al. 2012). Relative estimates, though, can be made to reasonably high accuracy (e.g., Tremonti et al. 2004), as long as a consistent estimator is being used. As a consequence, the simplified fits of N06 should be avoided when sufficient emission lines are available to make a more reliable and direct estimate of the metallicity (López-Sánchez & Esteban 2010).

Furthermore, the use of the N_2 parameter to derive Z is not valid in the high metallicity regime. This can

be appreciated in Fig. 1b, which shows the relationship between the T04-derived Z and the N_2 parameter. It is clear that the N_2 index saturates for $12+\log(O/H) \gtrsim 8.8$.

This agrees with the result found by Yin et al (2007), who concluded that empirical calibrations involving the N_2 parameter are only valid for galaxies with $12+\log(O/H) < 8.5$ –when comparing with T_e -based abundances; this limit should be 8.7–8.9 when Z has been derived using photoionization models, as in our case, see López-Sánchez et al. (2012)–.

The use of the N_2 method in a metallicity regime for which it is not valid should therefore be avoided. Even when this method is being averaged with another (e.g. R_{23} , Mannucci et al. 2010; Nagao, Maiolino & Marconi 2006), this can still drastically affect the resulting dependencies between SFR, M_\star , and Z . For example, Figure 1 of Yates et al. (2012) shows the SFR dependence on the M_\star – Z relation using the N06 and T04 metallicities. It can be appreciated from that figure that the dependence on SFR disappears in the high metallicity regime when the N06 metallicities are used. This lack of dependence can be explained by the saturation of the N_2 method in the high metallicity regime. On the other hand, when the T04 metallicities are used, this dependence is observed over the full metallicity range. To further support this observation, Lara-López et al. (in preparation) analyzed the same dependence using several combinations of metallicity (e.g., Pettini & Pagel 2004; Tremonti et al. 2004; Kewley & Dopita 2002) and SFR (e.g. Hopkins et al. 2003; Brinchmann et al. 2004) indicators, obtaining for all the possible combinations a strong dependence of SFR in the M_\star – Z relation for the full metallicity range.

4. A 3D analysis of the M_\star , Z , and SFR space

Here we aim to identify the most compact representation of the data distribution in the 3D space of M_\star , Z and SFR. We examine three methodologies: (i) fitting a plane to the 3D distribution using PCA, (ii) fitting a plane through regression (Lara-López et al. 2010a), and (iii) binning in SFR and M_\star to obtain the median metallicity in each bin (Mannucci et al. 2010).

As we want to perform a self-consistent comparison of the results, and because of the metallicity issues described in § 3, we do not use the Mannucci et al. (2010) method to derive Z , in the third approach mentioned above. Rather we test all the approaches self-

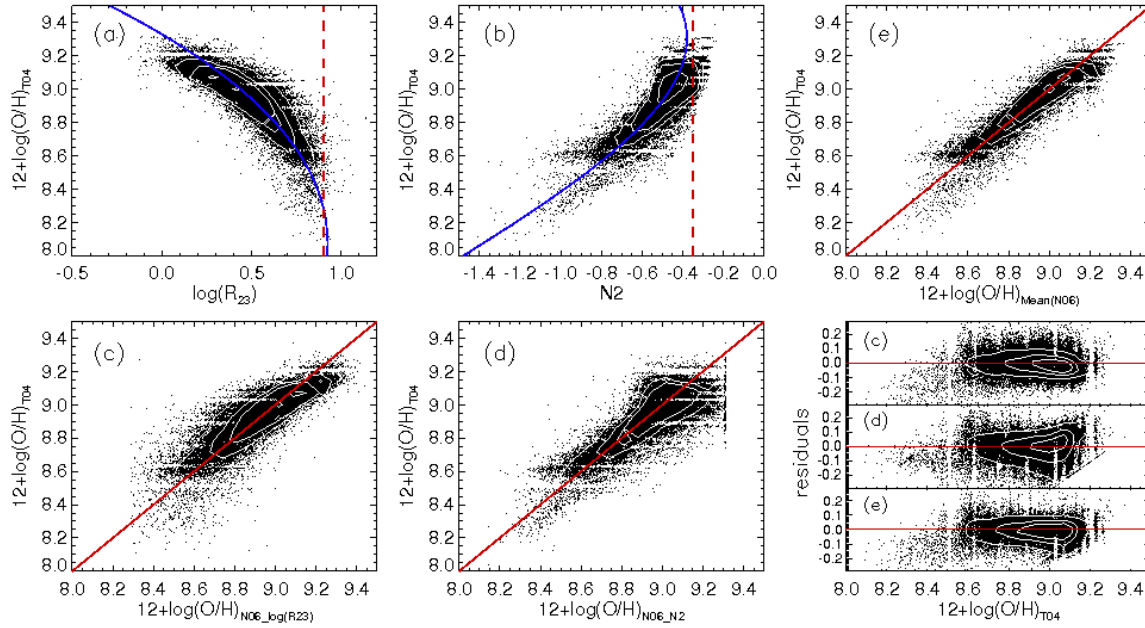


Fig. 1.— Panels *a* and *b* compare the $\log(R_{23})$ and N_2 parameters with the T04 metallicities, respectively. The blue line gives the N06 calibrations, while the red dashed line indicates the limit imposed by Mannucci et al. (2010) to their cubic fits. Panels *c* and *d* show the metallicity obtained by the N06 calibrations using the $\log(R_{23})$ and N_2 indexes vs. the T04 metallicities, respectively. Panel *e* shows the mean of the N06 metallicities when the difference between $\log(R_{23})$ and N_2 is less than 0.25 dex. The bottom right panel shows the residuals of panels *c*, *d* and *e*. We emphasize that the residuals show a scatter comparable with the intrinsic width of the $M - Z$ relation.

consistently using the MPA-JHU measurements detailed § 2.

Throughout this section we use the reduced chi-squared as a measure of goodness of fit, defined by the following equation:

$$\chi_{red}^2 = \frac{1}{\nu} \sum \frac{(O_i - E_i)^2}{\sigma^2} \quad (1)$$

where O and E are the observed data and model estimate, respectively. The value ν is the number of degrees of freedom given by $\nu = N - n - 1$, where N is the number of observations, and n is the number of fitted parameters. σ^2 is the variance of the observations defined by:

$$\sigma^2 = \frac{1}{N} \sum (O_i - \mu)^2 \quad (2)$$

where μ is the mean of the observed data.

4.1. PCA analysis

We performed a principal component analysis (PCA) to identify the underlying dimensionality of the three observables. PCA is a mathematical procedure that converts a set of observations of possibly correlated variables into a set of uncorrelated variables called principal components. One of the goals of PCA is to reveal hidden structure in a dataset, as well as to reduce the dimensionality of the data. A high correlation between variables is an indicator of high redundancy in the data, while the most important and independent variables are those that account for the largest variance. Since high correlation is a mark of high redundancy, the principal components should have low or even zero correlation between them (Shlens 2009).

We find that the first two principal components account for 86% and 12% of the variance, which indicates that 98% of our data can be explained in a 2 dimensional space. As a result of the PCA procedure we obtain 3 eigenvectors, referred to here as comp1, comp2, comp3, which are expressed as the vector of coefficients of the three parameters, $x=M_\star$, $y=12+\log(O/H)$, and $z=\log(SFR)$. The first principal component indicates the direction of the highest variance, and is given by comp1 = (0.7140, 0.1679, 0.6796), while the second and third components account for the highest possible variance in orthogonal directions: comp2 = (0.5952, 0.3654, -0.7156), comp3 = (0.3686, -0.9155, 0.1609), see Fig. 2. It is important to note here that one of the weak points of PCA is that

it relies on the covariance matrix, which is less robust against outliers than other methods. We return to this point in the discussion below.

The plane obtained through PCA is given by:

$$\alpha [M_\star/M_\odot] + \beta [12 + \log(O/H)] + \gamma [\log(SFR)] = \delta \quad (3)$$

where $\alpha = 0.3686$, $\beta = -0.9155$, $\gamma = -0.1609$, $\delta = -4.5578$

Solving Eq. 3 in turn for each of M_\star , Z and SFR, as a function of the other two, we obtain the relations shown in Fig. 3. Although PCA gives an acceptable result in reproducing the M_\star with $\chi_{red}^2 = 0.34$, the metallicity and especially the SFR present a high χ_{red}^2 of 1.0 and 0.94, respectively, (Fig. 3). It can also be appreciated in Fig. 3 that the fit provided by PCA is less effective at capturing the shape of the distribution, in particular at the low mass and metallicity end. This is a consequence of the covariance matrix sensitivity to outlying data points. This sensitivity effectively pushes the principal components to try to represent extremes in the data that are not representative of the bulk of the measurements.

4.2. Regression

We next explore regression to represent our 3D data distribution. Regression aims to explain one variable in terms of the others, and uses robust methods that are less affected by outlying observations. Using regression we first fit a plane to M_\star as a function of SFR and Z , obtaining:

$$\log(M_\star/M_\odot) = \alpha_m [12 + \log(O/H)] + \beta_m [\log(SFR)] + \gamma_m \quad (4)$$

where $\alpha_m = 1.3824$, $\beta_m = 0.5992$, $\gamma_m = -2.5729$.

The plane derived in this way was called the Fundamental Plane (FP) by Lara-López et al. (2010a) and is represented in Fig. 2 in blue. The difference between Eq. 4 and that presented in Lara-López et al. (2010a) is that here we are including galaxies at higher redshifts, up to $z \sim 0.35$. This improves the sampling of the high mass galaxy population, thus slightly changing the orientation of the plane. A detailed discussion of redshift and mass completeness will be given in Lara-López et al. (in preparation).

Although the FP (Eq. 4) is defined to minimize the variance in the estimate of M_\star , we also tested it to estimate the metallicity and SFR of galaxies by rearranging Eq. 4 to solve for the other variables (Fig. 4).

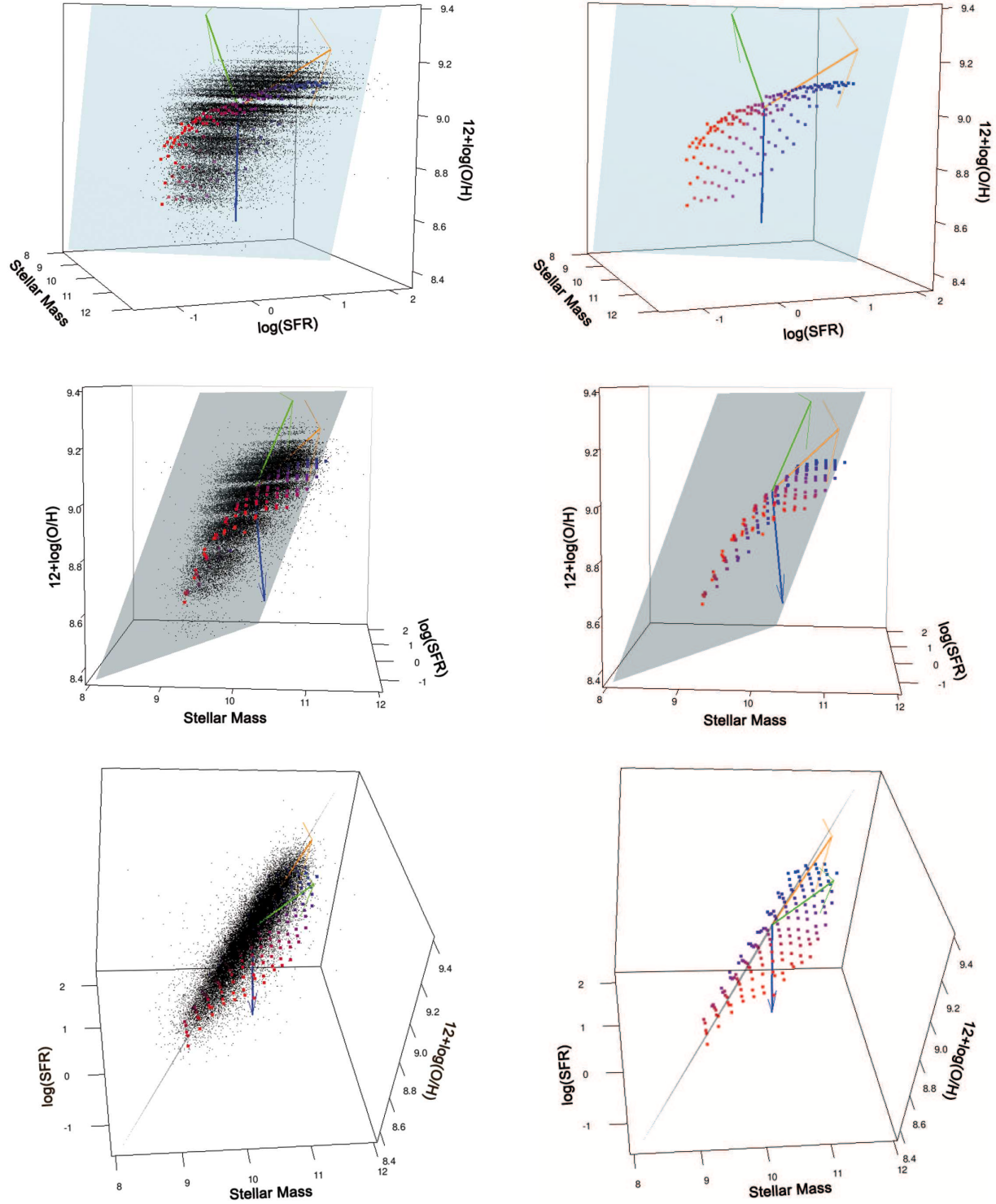


Fig. 2.— Different orientations of the 3D space formed by M_\star , SFR, and Z . The metallicity is always kept on the vertical axis, the cube is just moved rightwards, then tilted forward slightly, in going from the top to the bottom panels. The left panels show our newly derived FP (shaded), while the colored square points show the median metallicity taken in bins of SFR and M_\star (as for the FMR, Mannucci et al. 2010). Square points are color-coded from low (red) to high (blue) SFR. Black data points are the full sample (§ 2). The vectors show the first PCA component in yellow, the second in green, and the third in blue. The right panels show the same orientation and information as the left panels but omitting the underlying sample of SDSS galaxies. Upper panels show a face-on view of the FP, middle panels show an angle close to the $M_\star - Z$ relation, while bottom panels show the FP at the lowest dispersion face.

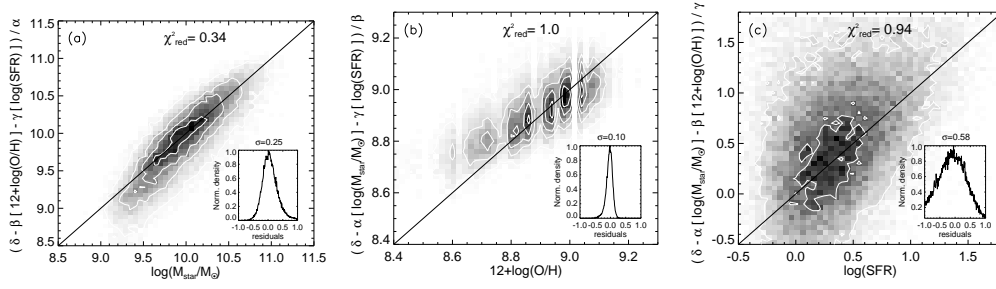


Fig. 3.— PCA plane. From left to right, M_* , metallicity and SFR estimated through PCA. The solid line shows the one to one relation, and the inset shows the histogram of the residuals.

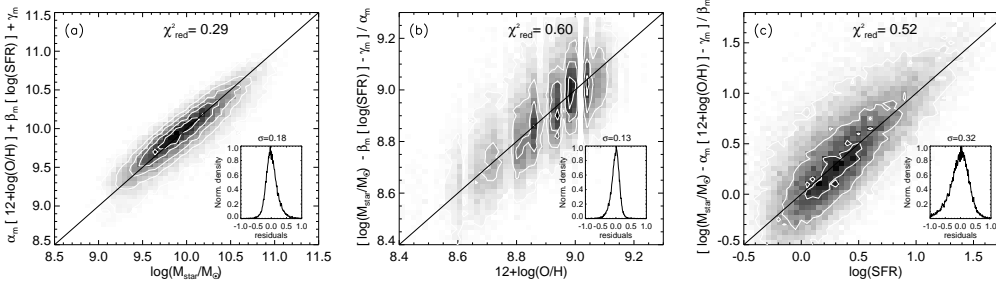


Fig. 4.— Plane fitted to M_* using regression, $M_* = f(Z, \text{SFR})$, the FP approach of Lara-López et al. (2010a).

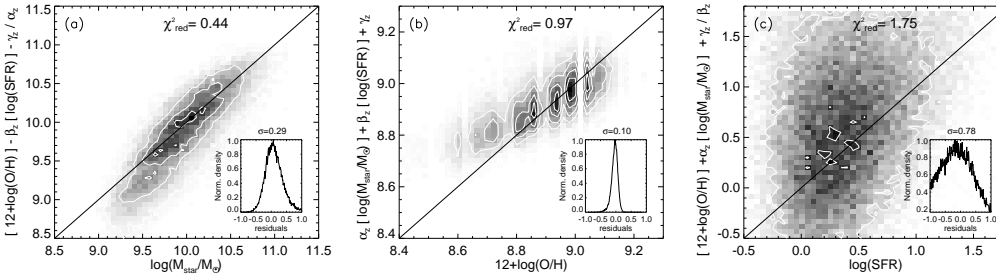


Fig. 5.— Plane fitted to Z using regression, $Z = f(M_*, \text{SFR})$.

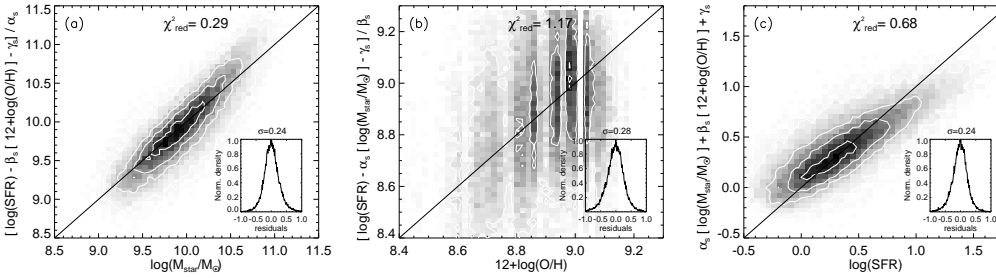


Fig. 6.— Plane fitted to SFR using regression, $\text{SFR} = f(Z, M_*)$.

Fit	M_\star		Z		SFR	
	χ_{red}^2	σ	χ_{red}^2	σ	χ_{red}^2	σ
PCA	0.34	0.25	1.0	0.1	0.94	0.58
$M_\star = f(Z, \text{SFR})$ (FP)	0.29	0.18	0.60	0.13	0.52	0.32
$Z = f(M_\star, \text{SFR})$	0.44	0.29	0.97	0.1	1.75	0.78
$\text{SFR} = f(M_\star, Z)$	0.29	0.24	1.17	0.28	0.68	0.24
FMR	–	–	0.88	0.1	–	–

Table 1: Summary of χ_{red}^2 and σ of the residuals for M_\star , metallicity and SFR using planes derived by PCA, regression, and the FMR.

Estimating χ_{red}^2 as before, we see that the metallicity obtained through the FP gives $\chi_{red}^2 = 0.60$. Fig. 4b also demonstrates that the estimate of Z follows a linear behaviour over the full metallicity range. We also tested this FP to estimate the SFR of galaxies, shown in Fig. 4c, obtaining $\chi_{red}^2 = 0.52$, which is an improvement compared to the $\chi_{red}^2 = 0.94$ obtained for the SFR through PCA. Again, Fig. 4c shows a more linear relationship through the full range of SFR compared to PCA.

To test whether the choice of M_\star as the dependent variable in Eq. 4 is the optimal approach, we also analyzed the two other possible planes, fitting in turn to Z and SFR (Fig. 5 and Fig. 6).

Fitting a plane to Z as a function of SFR and M_\star , we obtain:

$$12 + \log(\text{O}/\text{H}) = \alpha_z [M_\star/M_\odot] + \beta_z [\log(\text{SFR})] + \gamma_z \quad (5)$$

where $\alpha_z = 0.3504$, $\beta_z = -0.1289$, $\gamma_z = 5.4882$, giving $\chi_{red}^2 = 0.97$ for the metallicity. To compare this plane with the others, we again also estimate M_\star and SFR by rearranging Eq. 5. We find $\chi_{red}^2 = 0.44$ for M_\star , and $\chi_{red}^2 = 1.75$ for SFR (Fig. 5).

We also fit SFR as a function of Z and M_\star to obtain:

$$\log(\text{SFR}) = \alpha_s [M_\star/M_\odot] + \beta_s [12 + \log(\text{O}/\text{H})] + \gamma_s \quad (6)$$

where $\alpha_s = 0.9924$, $\beta_s = -0.8511$, $\gamma_s = -1.9167$, and giving $\chi_{red}^2 = 0.68$ for the SFR. Again, we estimate M_\star and Z by rearranging Eq. 6. This gives $\chi_{red}^2 = 0.29$ for M_\star , and $\chi_{red}^2 = 1.17$ for Z .

4.3. Binning data

Following Mannucci et al. (2010), we generated a grid of 0.11 dex in $\log(\text{SFR})$, and 0.15 dex in M_\star and estimated the median metallicity in every square of the

grid. The resulting values are shown in Fig. 2. This figure only shows median values for those bins containing at least 50 galaxies. It can be seen that, despite the different metallicity estimator, we can reproduce the shape of the surface obtained by Mannucci et al. (2010). By showing the underlying data used in deriving this surface (left panels of Fig. 2), it becomes clear that the curvature in this surface is not representative of the actual data distribution.

To compare how accurately the FMR can reproduce metallicity we follow Yates et al. (2012), since they use the same SDSS measurements of metallicity, SFR, and M_\star as in the current work. We used Eq. (1) of Yates et al. (2012) to estimate $\mu_\alpha = \log(M_\star) - \alpha \log(\text{SFR})$, with $\alpha = 0.19$. We then estimated metallicity using Eq. (2) of the same paper, given by $12 + \log(\text{O}/\text{H}) = 43.4476 - 12.1927x + 1.37284x^2 - 0.0498564x^3$, with $x = \mu_\alpha$.

The comparison between the metallicity obtained through the FMR and the original value gives $\chi_{red}^2 = 0.88$. (Fig. 7).

4.4. Results

The summary in Table 1 indicates that the plane that best minimizes χ_{red}^2 for M_\star , Z , and SFR, is the plane corresponding to the regression on M_\star , the FP. Also, this plane is the only one that maintains linearity over the full range of values between the observed and the estimated values in each projection. This further confirms that a plane is a good representation of the whole 3D distribution.

Empirically, we can imagine that the M_\star – Z , M_\star –SFR, and Z –SFR relationships are the faces of this 3D distribution. The relationship that shows the highest dispersion (§ 5) is the Z –SFR (Lara-López et al. 2010a,b), which means that this relation is close to the face-on orientation of the 3D distribution (top panels

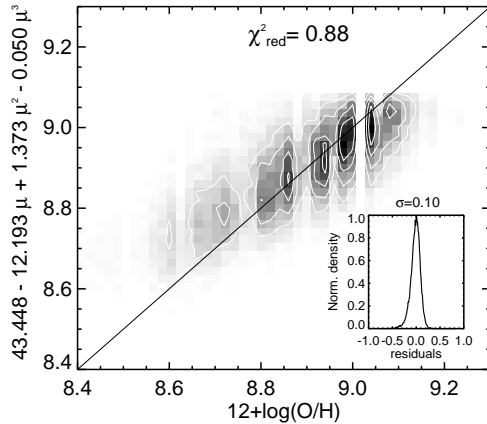


Fig. 7.— Metallicity of SDSS galaxies compared to the metallicity estimated through the FMR. The solid line shows the one to one relation.

of Fig. 2).

While M_\star correlates with both SFR and metallicity (the well known M_\star – Z and M_\star –SFR relationships), the SFR does not strongly correlate with metallicity. Therefore, SFR and Z are the least correlated variables, and a linear combination of these two parameters is enough to explain all three variables.

5. The Z –SFR relation

We turn now to a discussion of the Z –SFR relation, and the impact of taking the medians of the data in bins defined in different order. We have just described the Z –SFR relation as being close to the face on view of the FP, and the correlation between Z and SFR is not intrinsically tight. In consequence, any fit will suffer from a high degree of intrinsic scatter in the data around the fit.

This statement can be quantified using the Pearson correlation coefficient, which is a test of correlation between two variables, and is defined as $c = (\sigma_a^2 \times \sigma_b^2) / (\sigma_a \times \sigma_b)$. The quantity c is a scalar in the interval $[-1.0, 1.0]$, where -1.0 and 1.0 indicate a negative or positive perfect fit, respectively, while values close to 0.0 would indicate a poor correlation. Applying this test to our relationships, we obtain $c = 0.72$ for the M_\star – Z relation, $c = 0.76$ for the M_\star –SFR relation, and $c = 0.48$ for the Z –SFR relation. Therefore, of our three relationships, the Z –SFR is the relation that shows the highest dispersion.

Taking Z as the key quantity, Mannucci et al. (2010) studied the M_\star dependence of the Z –SFR relation by first taking bins of M_\star and then estimating the median Z in SFR bins. This procedure can be thought as the projection of the FMR on the M_\star –SFR face of the 3D distribution. The result is shown in Fig. 8a. Although this relation has a high scatter, there is a tendency for the SFR generally to increase with Z . Fig. 8a shows, however, that binning as a function of mass reveals a more subtle effect, with the resulting median values showing opposing trends depending on the mass selected. While the median metallicity increases with increasing SFR for $\log(M_\star) \gtrsim 10.5$, it decreases with increasing SFR for $\log(M_\star) \lesssim 10.5$ (Fig. 8a).

The mass dependence of the Z –SFR relation can alternatively be explored by binning in a different order. For every M_\star bin, we can estimate the median SFR in metallicity bins. Fig. 8b clearly shows that this binning order gives an apparently different result. For every M_\star , the SFR is almost flat or slightly increasing with the metallicity. There is a crucial distinction to be made here, which is one of correct interpretation. In the former case, the median *metallicity* has been estimated for a given M_\star and SFR, while in the latter, the median *SFR* has been estimated for a given M_\star and Z . It is true to say from the former approach that, at a given mass, as SFR increases, the median *metallicity* either increases (high-mass) or decreases (low-mass). It is also true to say from the latter approach that, at a given mass, as metallicity increases, the median *SFR* either increases (high-mass) or stays relatively constant (low-mass). These statements are not inconsistent with each other. The confusion arises when trying to impose an interpretation that is inconsistent with the motivation for binning in a particular order. It would *not* be true to conclude from Fig. 8a, for example, that the median *SFR* decreases with increasing metallicity (low-mass) or increases with increasing metallicity (high-mass), since the median *SFR* has not been calculated here.

We emphasize here that when fitting a relation to a distribution of data, the full dataset should be used. Fitting only to a representation of the full dataset that is derived from the median values of one parameter in bins of the others will clearly result in different surfaces being derived, and will depend on the choice of the parameter for which the median is estimated.

A detailed study to tease apart the complex interplay of M_\star , Z , SFR, and SSFR that exploits this use of

different binning order for all the relationships will be presented in Lara-López et al. (in preparation).

6. Discussion and conclusions

We have discussed the metallicity estimator of N06 showing that it is not accurate at high metallicities due to the saturation of the N2 method in the high metallicity regime. The use of a reliable metallicity estimator is crucial when analyzing the SFR dependence of the M_\star – Z relation as shown in Yates et al. (2012).

The emission-line galaxy spectra from SDSS are high quality, and measurements for many emission lines are available, making it possible to determine the gas-phase metallicity more robustly by applying techniques which consider the ionization degree of the gas. Examples of these methods are McGaugh (1991); Kewley & Dopita (2002); Kobulnicky & Kewley (2004) and Tremonti et al. (2004) (which are based on photoionization models) and Pilyugin (2001a,b); Pilyugin & Thuan (2005) and Pilyugin et al. (2010) (which rely on datasets for which Z is known using the T_e method).

We analyzed the 3D distribution of M_\star , Z , and SFR using three different approaches: (i) fitting a plane using PCA, (ii) fitting a plane through regression (Lara-López et al. 2010a), and (iii) binning in SFR and M_\star to obtain the median metallicity of each bin (Mannucci et al. 2010). For the five methods used, we estimated the χ^2_{red} as a measure of goodness of fit (Table 1). We find that the best representation of the data is the plane defined by regression on M_\star .

We compare the Mannucci et al. (2010) surface (the FMR) and the Lara-López et al. (2010a) Fundamental Plane (FP), and demonstrate that the FMR proposed by Mannucci et al. (2010) does not provide the best representation of the data. While PCA also does not provide the best representation, it does demonstrate that the 3D distribution can be adequately represented in 2 dimensions (a plane). The M_\star – Z , M_\star –SFR, and Z –SFR relationships are then projections of this plane. We highlight here that using the Mannucci et al. (2010) FMR to estimate metallicities from M_\star and SFR can result in large errors, up to ~ 0.2 dex in oxygen abundance (Fig. 7), being larger at high metallicities as a consequence of the inappropriate use of the N_2 index (López-Sánchez & Esteban 2010; López-Sánchez et al. 2012).

We also highlight that the plane found by the regression fit on M_\star is not developed as a new technique for stellar mass estimation. Rather, this approach is pri-

marily aimed at identifying the most concise representation of M_\star , Z and SFR in order to facilitate more detailed exploration of the interplay between these properties of galaxies. Nevertheless, in cases when more robust techniques to estimate the stellar mass (e.g., Taylor et al. 2011) are not available, the use of the FP could be used to estimate the stellar mass, being careful to take into account the metallicity and SFR uncertainties.

Our analysis of the Z –SFR relation with the two approaches toward binning the data highlights a crucial need for caution in interpretation when exploring distributions represented as median values. The inappropriate interpretation of such results will lead to apparently contradictory conclusions, depending on the binning order used. Furthermore, presenting medians in bins as a three-dimensional distribution will lead to differing representations of the data, depending on the binning order chosen.

The SFR of a galaxy relates to the amount of gas currently being converted into stars, and correlates with the current mass in stars, while metallicity is a measure of the number of times that the gas has been reprocessed by stars, and also correlates with the current mass in stars in a galaxy. The fact that we can represent M_\star as a linear combination of SFR and metallicity suggests that the stellar mass of a galaxy can be thought as the rate at which a galaxy is currently forming stars (SFR), plus a measure of the star formation history, which is here represented by the metallicity (Z), corresponding to the amount of reprocessing of the gas by past stellar generations. The SF history and current SFR of a galaxy are closely linked to M_\star .

To further support this result, Lara-López et al. (2010a) demonstrated that high redshift data agree with this FP, with a lack of evolution in the FP up to $z \sim 3.5$. This is explained by SFR and metallicity evolving in opposite directions: while high redshift galaxies show higher SFR than local galaxies, they also present lower metallicity abundances than local systems. These differences make it possible to recover, independently of redshift, the M_\star of a galaxy as a combination of SFR and Z .

We warmly thank the referee for extensive comments that have led to significant improvement in this paper. We thank Mercedes Mollá, César Esteban, Alessandro Ederoclite, and Dámaso E. Chicharro-Martínez for useful comments. M. A. Lara-López thanks the “Summer School in Statistics for Astronomers”,

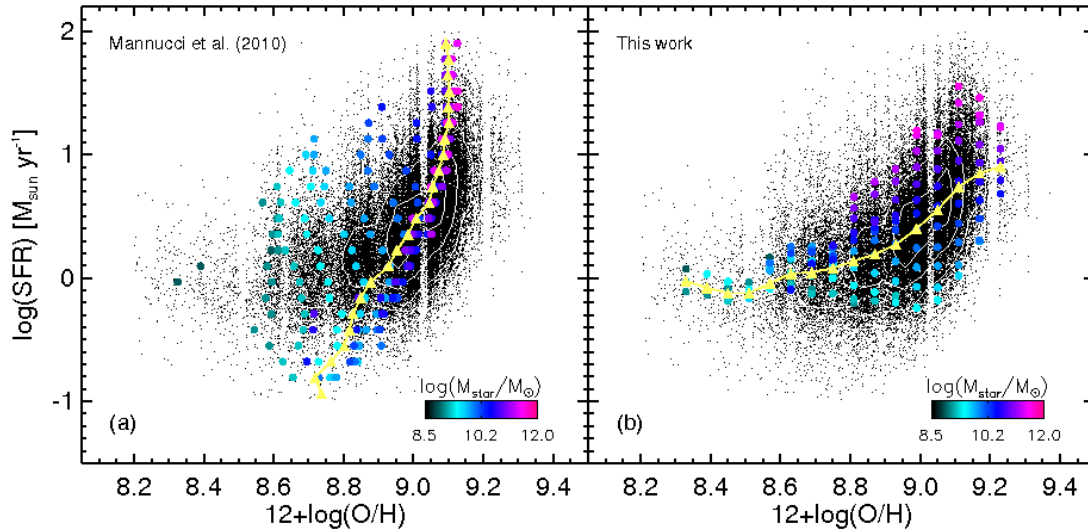


Fig. 8.— (a) Filled coloured circles correspond to the Z -SFR relation binned following Mannucci et al. (2010), who took bins in $\log(\text{SFR})$ and estimated the median $12+\log(\text{O}/\text{H})$ per bin for different ranges of $\log(M_\star/M_\odot)$, as shown in the color bar. (b) The same data, but now taking bins of $12+\log(\text{O}/\text{H})$ and estimating the median $\log(\text{SFR})$ using the same mass ranges as in (a). Black dots and white contours correspond to the SDSS sample. Yellow triangles show (a) the median metallicity in SFR bins, or (b) the median SFR in metallicity bins, considering the whole sample.

Center for Astrostatistics, PennState, especially to Arnab Chakraborty for invaluable tutorials on “R” and PCA.

The work uses Sloan Digital Sky Survey (SDSS) data. Funding for the SDSS and SDSS-II was provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS was managed by the Astrophysical Research Consortium for the Participating Institutions. This research has made use of the SAO/NASA Astrophysics Data System Bibliographic Services (ADS).

REFERENCES

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, *ApJS*, 182, 543
- Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., et al. 2007, *ApJs*, 172, 634
- Baldwin J., Phillips M., Terlevich R., 1981, *PASP*, 93, 5 (BPT)
- Bresolin, F., Gieren, W., Kudritzki, R-P., Pietrzyński, G., Urbaneja, M.A. & Carraro, G. 2009, *ApJ*, 700, 309
- Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, *MNRAS*, 351, 1151
- Charlot, S., & Longhetti, M. 2001, *MNRAS*, 323, 887
- Davé, R. & Oppenheimer, B.D. 2007, *MNRAS*, 374, 427
- De Lucia, G., Kauffmann, G. & White, S.D.M. 2004, *MNRAS*, 374, 323
- De Rossi, M.E., Tissera, P.B., & Scannapieco, C. 2006, *MNRAS*, 374, 323
- Ellison, S. L., Patton, D. R., Simard, L., et al. 2008, *ApJL*, 672, L107
- Esteban, C., Bresolin, F., Peimbert, M., García-Rojas, J., Peimbert, A. & Mesa-Delgado, A. 2009, *ApJ*, 700, 654
- García-Rojas, J. & Esteban, C., 2007, *ApJ*, 670, 457
- Hopkins, A. M., Miller, C. J., Nichol, R. C., et al. 2003, *ApJ*, 599, 971

- Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, MNRAS, 346, 1055
- Kewley, L.J. & Dopita, M.A. 2002, ApJS, 142, 35
- Kewley, L.J., & Ellison, S.E. 2008, ApJ, 681, 1183
- Kewley, L. J., Jansen, R. A., & Geller, M. J. 2005, PASP, 117, 227
- Kobulnicky H. A. & Kewley L. J. 2004, ApJ, 617, 240
- Lara-López, M. A., Cepa, J., Bongiovanni, A., et al. 2010a, A&A, 521, L53
- Lara-López, M. A., Bongiovanni, A., Cepa, J., et al. 2010b, A&A, 519, A31
- Lequeux, J., Peimbert, M., Rayo, J. F., Serrano, A., & Torres-Peimbert, S. 1979, A&A, 80, 155
- Liang, Y. C., Yin, S. Y., Hammer, F., et al. 2006, ApJ, 652, 257
- López-Sánchez, Á.R. & Esteban, C. 2009, A&A, 508, 615
- López-Sánchez, Á.R. & Esteban, C. 2010, A&A, 517, 85
- López-Sánchez, Á.R., Mesa-Delgado, A., López-Martín, L & Esteban, C. 2011, MNRAS, 411, 2076
- López-Sánchez, Á.R., Esteban, C., García-Rojas, J., Peimbert, M. & Rodríguez, M. 2007, ApJ, 656, 168
- López-Sánchez, Á.R., Dopita, M.A., Kewley, L.J.-Zahid, H.J., Nicholls, D.C. & Scharwächter, J. 2012, MNRAS, in press
- Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115
- Martínez-Serrano, F. J., Serna, A., Domínguez-Tenreiro, R., & Mollá, M. 2008, MNRAS, 388, 39
- McGaugh, S.S. 1991, ApJ, 380, 140
- Mesa-Delgado, A., & Esteban, C. 2010, MNRAS, 405, 2651
- Mollá M., Díaz A. I., 2005, MNRAS, 358, 521
- Moustakas, J., Kennicutt, R.C., Jr., Tremonti, C. A., Dale, D. A., Smith, J.-D. T. & Calzetti, D. 2010, ApJS, 190, 233
- Nagao, T., Maiolino, R. & Marconi, A. 2006, A&A, 459, 85
- Peimbert, M., Peimbert, A., Esteban, C., et al. 2007, RMxAC, 29, 72
- Pettini, M. & Pagel, B.E.J. 2004, MNRAS, 348, 59
- Pilyugin, L.S. 2001a, A&A, 369, 594
- Pilyugin, L.S. 2001b, A&A, 374, 412
- Pilyugin, L.S. & Thuan, T.X. 2005, ApJ, 631, 231
- Pilyugin, L.S., Víchez, J.M. & Thuan, T.X. 2010, ApJ, 720, 1738
- Shlens, J., 2009, *A Tutorial on Principal Component Analysis, Version 3.1*, New York University/Systems Neurobiology Laboratory, Salk Institute for Biological Studies
- Strauss, M. A., et al. 2002, AJ, 124, 1810
- Taylor, E. N., Hopkins, A. M., Baldry, I. K., et al. 2011, MNRAS, 418, 1587
- Tissera, P.B., De Rossi, M.E., & Scannapieco, C. 2005, MNRAS, 364, L38
- Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898
- Yates, R. M., Kauffmann, G., & Guo, Q. 2012, MNRAS, 422, 215
- Yin, S.Y., Liang, Y.C., Hammer, F., Brinchmann, J., Zhang, B., Deng, L.C. & Flores, H., 2007, A&A, 462, 535